Vol.2022-GN-117 No.6
Vol.2022-SPT-47 No.6
2022/5/26

グループ対話の印象を正しく伝える映像要約手法

高山千尋[1]　永徳真一郎[1]　二瓶芙巳雄[1]　石井亮[1]
中野有紀子[2]　深山篤[1]　中村高雄[1]

概要：グループ対話の内容，およびその場の雰囲気を効率的に伝達するための要約手法を提案する．提案手法では，対話映像を対象に，重要発言・談話行為の情報に基づいて，発言箇所を抽出し要約映像を生成する．提案手法による要約結果の評価，オリジナル映像からの雰囲気の印象変化について，被験者による評価実験を行った．ランダムに発言抜き出す要約，専門家による要約と比較したところ，提案手法は，ランダム要約より優れていること，雰囲気伝達の観点で人手による要約と大きな差がない可能性が示唆された．また，提案手法は，重要発言のみを抜き出した要約と比較して，雰囲気の伝達に優れていること，雰囲気の変化が少ないことが示された．

キーワード：要約技術，抽出要約，多人数対話

Video Summarization Technique to Correctly Convey Impressions of Group Dialogue

CHIHIRO TAKAYAMA[†1]　SHINICHIRO EITOKU[†1]　FUMIO NIHEI[†1]
RYO ISHII[†1]　YUKIKO NAKANO[†2]　ATSUSHI FUKAYAMA[†1]
TAKAO NAKAMURA[†1]

1. はじめに

近年，映像機器が安価で入手しやすくなり，また映像を保存・編集・配信するサービスの普及が進んでいる．それに伴って，記録としての映像の利用が盛んになっている．加えて，COVID-19 のパンデミックにより，遠隔でのオンライン会議が一般的になり，対話の映像がより多く取り扱われるようになってきた．このような情勢から，記録された映像を効率的に視聴するニーズが高まってきている．

このような，大量の記録を人が効率的に振り返るための技術として，発話音声の書き起こし技術，またそれらを要約した議事録作成技術が研究されてきた[1]．しかし，これらの文字を使った要約では内容は伝えられたとしても，意見に対する参加者の賛成や反対の程度，会議全体の活性度などの対話の様子は十分に伝えられない．

文字により表現された情報と比較して，映像の視聴には時間がかかる．再生速度を変更できたとしても，内容を理解できるのはせいぜい 2〜4 倍速程度までが限界である．そのため，映像を効果的に要約する技術も研究されている[2]．しかしながら，これらの技術は内容の伝達のみを対象としており，文字と比較して映像がより得意とする，対話の雰囲気に関する情報を十分に伝えることができない．

本研究では，対話における内容だけでなく，雰囲気も正確に伝えるための要約技術の確立を目指している．なお，ここでの「雰囲気」とは，対話参加メンバおよび彼らによる対話全体に対する印象を指しており，評価のため経営・組織学における組織の集団雰囲気指標[3]を指標に採用した．そのうえで，同一ユーザによる要約前の対話映像に対する本指標の評価値と，要約後の対話映像に対する評価値とが一致することをもって「雰囲気」を正しく伝えられる要約である，と定義している．

提案手法は，元となる対話映像から，特定の発言箇所を抽出し繋ぎ合わせることで，要約映像を出力する．抽出箇所の選択については，意見を示す重要発言に加えて，その発言の前に発生する質問や議題提案などの問いかけの発言，さらにその発言の後に発生する賛同や反対などの反応の発言を加える方式を採用している．

この(1)提案手法について，3 対話を対象に，(2)ランダムに発言を抽出した要約と，(3)専門家による人手での要約とで被験者による比較評価を行った．その結果，(1)提案手法は，(2)ランダム要約より優れている点，雰囲気の伝達においては(3)人手による要約と遜色がない可能性が示唆された．

次に，4 対話に対して，(1)提案手法と，(4)重要発言のみ，(5)重要発言＋反応発言，(6)重要発言＋問いかけ発言の 4 つの要約手法を比較した結果，提案手法(1)が他の手法と比較して，より雰囲気を正確に反映している評価となった．

以降，2 章では関連研究について述べ，3 章で提案手法，4 章で評価実験について述べ，5，6 章でその結果および考察を述べる．

2. 関連研究

2.1 映像要約技術

要約手法は文章や映像など，様々なモーダル分野で研究が行われている．これらは，元のデータの一部を抽出して要約を生成する抽出要約（extraction summarization）と，元

Vol.2022-GN-117 No.6
Vol.2022-SPT-47 No.6
2022/5/26

のデータの内容を元に，新たにより短い記述を生成する抽象化要約（abstraction summarization）に分類される．映像においても，これらの手法が研究されている[2].

2.1.1 抽出要約

映像の抽出要約は，ニュースやスポーツ，ドラマなどのテレビ番組や，ホームビデオなどを対象とした研究が盛んにおこなわれている[4]．実際の映像を再利用するため，後述の抽象化要約と比較して真実性が高いとされる．この研究分野で盛んに用いられているデータセットは，SumMe[5]と TvSum[6]である．これらは，YouTube などにアップロードされたホームビデオ（SumMe は 1 分～6 分程度，TvSum は 15 分～25 分程度）のビデオクリップに対して，複数名の作業者で，要約映像として残す箇所をアノテーションしたデータセットである．これらデータセットを用い，メディア処理と機械学習の技術を用いて，映像や音声に含まれる特徴量から要約箇所を推定するタスクが広く行われている．しかしながら，これらのデータセットには対話の映像が含まれておらず，対話の抽出要約を対象とした研究は行われていない.

2.1.2 抽象化要約

映像を対象とした抽象化要約とは，元の映像の特徴量をもとに，映像の内容・特徴の記述文を生成する技術である[2]．映像とその説明文のデータセットを元に，機械学習における敵対的生成ネットワークなどの技術を用いて，新たに文章を生成する．これらは，主に映像の検索性を高めるための簡易な説明文の生成を目的としている.

本研究が対象とする対話の要約においては，対話の内容がわかる程度の長い文章と，正確な発言の記述（抜き出し）が必要であり，現時点の抽象化要約では，対話の発言内容や結論が変化してしまう点が指摘されており[7]，本研究が目指す用途には合わない.

2.2 議事録要約技術

対話の書き起こしに対する抽出・抽象化要約技術も検討されている[1]．この議事録要約技術では，会議などの複数名が行う対話のやり取りを発言ごと・話者ごとに書き起こし，その文字情報全体，あるいは単語ごとの特徴に基づいて，特定の発話を抜き出す手法や，概要を説明する文字列を生成する手法が提案されている.

この技術は，対話の論点や結論を把握するには効率が良いが，発言の書き起こしを対象としているため，発言の前後での参加者の言外の反応や，対話全体に対する印象や雰囲気，話者同士の関係などを詳しく把握することは難しい.

2.3 対話状況認識

上述の対話内容，および対話の雰囲気の理解を助けるため，発言の意図や，対話全体の雰囲気などを推定する手法の研究が進められている.

近藤ら[8]は，対話に介入するロボットの開発を目的として，対話の特徴量から，対話中の雰囲気が「良い・中間・悪い」と変化する様子を推定・評価する技術を提案している．しかしながら，雰囲気の尺度として，「良い・中間・悪い」の 3 分類だけでは，本研究が目指す対話の雰囲気の概要を把握するには不十分であると考える.

二瓶ら[9]は，発言に含まれる音声特徴量を用いて，重要発言（要約に必要と考えられる発言）を推定する手法，および対話映像を効率的に閲覧・内容把握するための議論ブラウザを提案している．被験者実験において，要約機能を持たない simple browser，書き起こし文章のみを表示する text-based browser，提案手法を実装した multimodal browser を比較し，理解の程度，閲覧時間，参加者の役割認識，使いやすさなど点から比較を行い，提案手法の有効性を示している．しかし，これらの手法は，内容把握には効果的であっても，対話の雰囲気の伝達に対しては不十分である．対話において重要と分類される発言は，情報提供目的・意見表明であることが多い．重要発言のみを抽出要約した場合，参加者が各々意見表明を行う要約映像となってしまい，その意見に至るまでの話の流れ，意見に対する参加者の反応などをうまく伝えることができない.

3. 提案手法

3.1 対話映像要約手法の全体像

提案する対話映像要約手法は，映像データから重要発言・発言ごとの談話行為を推定する前半部と，推定結果と要約対象の対話映像から，抽出箇所を選択する後半部に分けられる（図 1）．特に，後半の要約アルゴリズムにおいて，対話内容に関する重要発言だけでなく，その発言の前後の流れの理解を助ける「問いかけ発言」や，重要発言に対する参加者の反応を示す「反応発言」を要約映像に加える処理を行う．以下で，それぞれの詳細を説明する.

3.2 重要発言・談話行為の推定

対話映像から抽出した顔画像情報，音響情報，および音

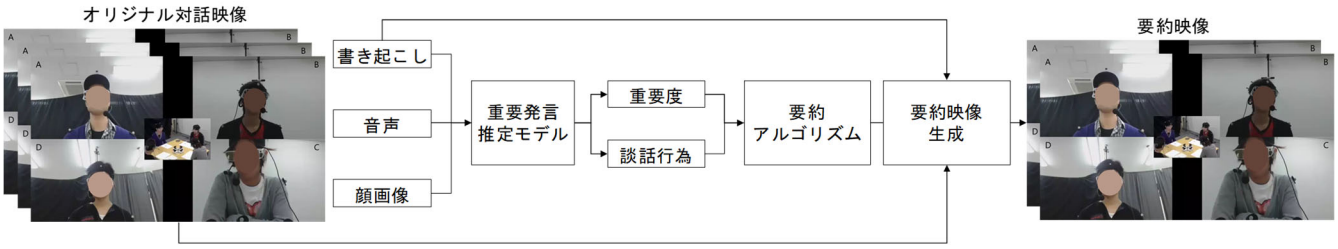

図 1 提案手法の全体像

声認識処理によって得られる書き起こし情報を使って，重要発言と談話行為を推定する．

重要発言の推定では，二瓶ら[9]が提案する重要発言推定モデルを利用する．音声のスペクトログラム，顔画像から得られる頭部動作のスペクトログラム，時間経過における音声のインテンシティの変化，時間経過における頭部姿勢の変化，そして発言者の言語情報の埋め込みベクトルを入力とする畳み込みニューラルネットを用いて，すべての発言の中から重要発言を推定する．ここで推定する重要発言とは，ラベル付けを行う 5 名の作業者が「議論要約に含めるべき発言を選択すること」という教示に従って， 5 名全員が重要と判断した発言のことである．

談話行為（Dialogue act）は，情報取得のための質問や話者関係をよくするための感謝など，発話ごとの発言者の意図を示す分類であり，ISO 24617-2:2020 [10] などで定義されている．本研究においては，談話行為推定タスクで一般的に用いられる AMI プロジェクトで提案されている 15 分類を利用する[11]．これら分類についても，上述した 5 つのモーダル情報を埋め込みベクトルを入力とする畳み込みニューラルネットを用いて，推定する．

3.3 抽出箇所の特定と要約映像の作成

次に，上述した方法で推定された重要発言や，談話行為の情報を用いて，オリジナルの対話映像から要約映像に残す区間を選択する．2.3 節で述べたとおり，重要発言のみの要約では，主要な意見は抽出できても，その意見がなされた議論の流れや，意見に対する参加者の反応を表すことが難しい．

そこで重要発言と判定された発言に加えて，談話行為が，参加者に発言を促す「問いかけ発言」と，参加者の発言に対して賛成, 反対, あるいは単純な反応を返す「反応発言」に分類される発言を追加した要約を作成する. 具体的には，「問いかけ発言」として，談話行為における対話タスクを進めるための Suggest/Offer や, 情報取得を目的とした Elicit Inform/Elicit Assessment を加える. また「反応発言」として，頷きを表す Backchannel や，理解していることを話者に対して返す Comment-about-Understanding，意見に対する評価を返す Assess を加える．

要約映像では，上記で選択された各発言の開始時刻～終了時刻について，オリジナル対話映像から抜き出し，書き起こしの文字情報をテロップとして加え，一つの要約映像とする．発言が重なる場合は，それらを結合した区間をまとめて抽出する．

4. 実験

提案する要約手法における抽出アルゴリズムの妥当性を評価するため，被験者を用いた比較評価実験を 2 回行った．実験 1 では，提案手法と，後述するランダム要約と人手による要約とを比較した．実験 2 では，提案手法のバリエーションを作成し比較した．提案手法後半の要約アルゴリズムの評価を目的とするため，評価実験では重要発言や談話行為の推定処理は行わず，人手でアノテーションを付与した対話データ（前半の推定におけるモデルの学習に用いた教師データ）を用いた．以下で実験の詳細を述べる．

4.1 仮説

これまで，対話映像の要約手法においてベンチマークとなる手法は提案されていない．そこで，本実験においてベースラインとして(2)ランダムに発言を抽出した要約，最良の結果として(3)人間による要約を作成し，これらと(1)提案手法とを比較することで，被験者が各要約手法を見分けられるかを確認するとともに，提案手法の評価を行う．そこで実験 1 では，以下の 2 つの仮説を検証する．

H1-1: 提案手法は，ランダム要約より良い評価となる．

H1-2: 提案手法は，人手の要約未満の評価となる．

次に, 提案手法のポイントである, 重要発言だけでなく，「反応発言」や「問いかけ発言」を加えることについて検証を行う．実験 2 では，要約手法の複数のバリエーションに対して，提案手法の評価を比較検証する．

H2: 提案手法は，他の要約手法（重要発言のみ，重要発言＋反応発言，重要発言＋問いかけ発言）より，評価が良い．

最後に，提案手法が内容だけでなく，雰囲気を正しく伝えられることを確認するため，以下の仮説を実験 2 で検証する．

H3: 提案手法は，他の要約手法と比べて，雰囲気の変化が少ない．

4.2 評価対象データ

評価する対話データとして，MATRICS コーパスデータを対象とした[12][13]．このデータは，初対面の 4 者が，20 分程度で学園祭の出店計画を議論し，出店場所・出店内容を決定する様子を撮影したものである．このうち，重要発言や談話行為をアノテート済みの 6 対話を要約対象データとして利用する．

(1)提案手法の比較対象として, (2)ランダムに発言を抽出した要約映像と，(3)専門家による人手要約を作成した．(3)について，映像制作会社を通じて応募いただいた，少なくとも 10 年以上業務として演出・編集に携わっている専門家に，3 対話について，要約映像を制作いただいた．

以上より実験 1 においては，3 つの対話について，以下の要約バリエーションを作成し評価実験を行った．

(1) 重要発言と問いかけと反応を抽出（提案手法）
(2) ランダム要約
(3) 専門家による人手要約

次に，実験 2 にて重要発言以外の発言を追加する効果を評価するため，提案手法の比較対象として，以下の 4 つのバリエーションを 4 対話について作成し，比較した．

(1) 重要発言と問いかけと反応を抽出（提案手法）

(4) 重要発言のみ抽出
(5) 重要発言と反応発言を抽出
(6) 重要発言と問いかけ発言を抽出

それぞれの対話のオリジナルと要約映像の再生時間は表 1，表 2 のとおりである．

表 1 対象とする対話と要約後の再生時間

	対話 1	対話 2	対話 3
オリジナル	20:04	20:09	20:04
(1)提案手法	7:20	5:26	8:29
(2)ランダム要約	5:05	4:54	4:50
(3)人手要約	5:00	5:00	5:01

表 2 対象とする対話と要約後の再生時間

	対話 3	対話 4	対話 5	対話 6
オリジナル	20:04	20:04	20:07	20:04
(1)提案手法	8:29	5:30	4:39	6:39
(4)重要発言のみ	6:24	3:52	3:09	3:58
(5)重要発言＋反応	7:43	5:08	4:09	6:17
(6)重要発言＋問いかけ	7:11	4:16	3:39	4:23

4.3 評価指標

要約の良さを測る指標として，対話や会議を後から視聴するニーズを考慮して，以下の 5 つの観点を設定した．

- Q1: この要約映像は，優れた要約だ
- Q2: この要約映像は，内容を反映している
- Q3: この要約映像は，納得感が高い
- Q4: この要約映像は，対話の雰囲気を示している
- Q5: この要約映像は，参加者の関係を反映している

これらの各観点について，スライダーを使って 1〜100 の値を選んで回答させた（以下，評点と呼ぶ）．

参加メンバおよび対話の雰囲気の印象を測る指標として経営・組織学における組織の雰囲気を測る「集団雰囲気尺度」を指標として採用した[3]．対話映像中の参加メンバの印象について，以下の 10 項目について 9 件法で回答させた（以下，雰囲気と呼ぶ）．

- 友好的 - 非友好的
- 受容的 - 拒絶的
- 満足させる - 挫折させる
- 熱烈な - 熱のない
- 生産的な - 非生産的
- 暖かい - 冷たい
- 協力的 - 非協力的
- 支持的 - 敵対的
- 面白い - 退屈な
- 成功する - 成功しない

4.4 被験者

実験 1，2 ともに，インターネットサイトを介して，募集を行った．被験者は，男女を均等に，年齢は 20‐70 歳に分布するよう割付を行った．4.5 節で説明する信頼性を測る指標によって，分析対象の絞り込みを行った．

実験 1 では，被験者間実験として，有意水準 α=.05，検出力(1-β)=.08，小さい効果量 f=0.10 を検出可能なサンプルサイズ 787 名にデータ欠損や不正確な回答者などの余裕分を加えた 1,000 名を目標に募集を行い 1,099 名からの応募を得た．1,099 名分の回答のうち，4.5 節で述べる方法によって回答の信頼性が低いと判断された 420 名（38.2%）の被験者を分析から除外し，結果的に 679 名の被験者の回答を分析した（男性 426 名，女性 253 名．21〜70 歳，平均 52.9 歳，標準偏差 9.9 歳）．

実験 2 では，有意水準 α=.05，検出力(1-β)=.08，小さい効果量 f=0.10 を検出可能な被験者内実験として，サンプルサイズ 200 名にデータ欠損や不正確な回答者などの余裕分を加えた 250 名を目標に募集を行い，250 名の応募を受けた．20 名（8%）の被験者を実験 1 と同様に分析から除外し，結果的に 230 名の被験者からの回答を分析した（男性 115 名，女性 115 名．20〜73 歳，平均 43.0 歳，標準偏差 13.6 歳）．

4.5 評価手法

実験 1 では，被験者間実験計画でのオンライン評価実験を行った．表 1 に記載の 3 種の要約手法×3 種の対話について，被験者一人当たり 4 つをランダムに順次提示し評価させた（作業時間は 30 分程度）．この作業を被験者 679 名に対し行い，各要約動画について 121〜172 回の回答を収集した（合計 1,342 回答）．

実験 2 では，被験者内実験計画でのオンライン評価実験を行った．被験者は，実験内容の説明を受けた後，オリジナルの対話映像（20 分）を視聴し，その対話の参加者の雰囲気（10 項目）について回答した．次に，先に視聴したオリジナル対話映像に対応する 4 種類の要約映像（3 分〜8 分）を視聴し，4.3 節で述べた要約の良さを測る指標（5 項目）についてそれぞれ回答し，さらに参加者の雰囲気（10 項目）についてそれぞれ回答した．被験者は，要約映像視聴中は，比較のためにいつでもオリジナルの対話映像に戻って視聴することが可能であった．被験者は，1 つのオリジナル対話映像と 4 つの要約映像の評価を 1 セットとして，表 2 に記載の 4 対話からランダムに割り当てられた合計 2 セット（オリジナル対話映像 2 つ，要約映像 8 つの計 10 つ．作業時間は合計 2 時間程度）に対して評価を行った（合計 920 回答）．

実験 1，実験 2 ともに，被験者の回答の信頼性を測る指標として，要約の良さを測る質問の一つに「この質問には 100 と回答してください．」と記載した質問を含め，被験者は必ず 1 回はこの質問に回答するよう設定した．この質問

Vol.2022-GN-117 No.6
Vol.2022-SPT-47 No.6
2022/5/26

への回答が 100 ではない被験者は信頼性が低いとして，分析から除外した．

統計分析では，評点，雰囲気の回答値は正規分布しなかったため，二元配置分散分析（対話×要約手法）を行ったのち，ノンパラメトリック分析による多重比較を行った．

5. 結果

5.1 提案手法とランダム要約・人手による要約との比較

対話（対話 1，対話 2，対話 3）と要約手法（(1)提案手法，(2)ランダム要約，(3)人手要約）との二元配置分散分析の結果，すべての評点において要約手法の主効果が有意な差が認められた（Q1:優れた要約 $F(2, 1333)=10.08, p<.01, \eta^2=.015$, Q2:内容を反映 $F(2, 1333)=12.40, p<.01, \eta^2=.018$，Q3:納得度が高い $F(2, 1333)=10.24, p<.01, \eta^2=.015$，Q4:雰囲気を示している $F(2, 1333)=6.11, p<.01, \eta^2=.009$ Q5:話者関係を反映 $F(2, 1333)=6.10, p<.01, \eta^2=.009$）．また，Q3 については，対話の主効果で有意な差が認められた（Q3: $F(2, 1333)=3.73, p=0.024, \eta^2=.005$）．いずれの評点においても，交互効果の有意差は認められなかった．

有意差が認められた要約手法について，Steel-Dwass 法による多重比較を行った結果が，図 2 である．すべての評点において，提案手法＞ランダム要約の順で有意に評点が高くなり，Q2:内容の反映と Q3:納得度が高いにおいて手動要約＞提案手法の順で有意に評点が高い結果となった．

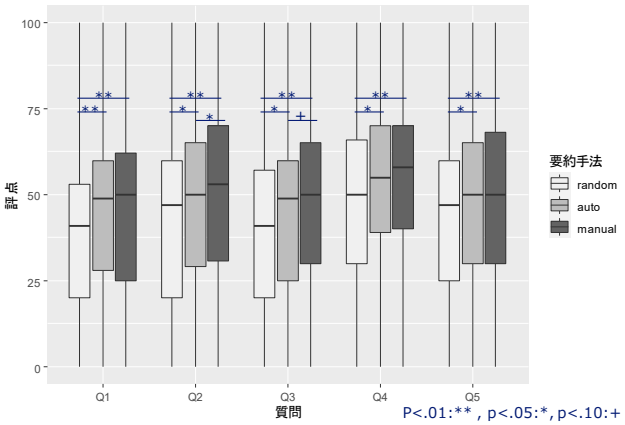

図 2 要約手法の評点の比較（(1)(2)(3)）
random:ランダム要約，auto:提案手法
manual:人手による要約

5.2 提案する要約手法のバリエーションの比較

対話（対話 3，対話 4，対話 5，対話 6）と，要約手法（(1)提案手法，(4)重要発言のみ，(5)重要発言と反応発言，(6)重要発言と問いかけ）との二元配置分散分析の結果，「Q4:雰囲気を示している」において対話の主効果（$F(3, 904)=4.98, p<.01, \eta^2=.016$），要約手法の主効果（$F(3, 904)=6.60, p<.01, \eta^2=.021$）で統計的に有意な差が見られた．また，「Q5:話者関係の反映」において，要約手法の主効果で統計的に有意な差が見られた（$F(3, 904)=2.65, p=.048, \eta^2=.009$）．交互効果については，いずれの評点においても有意差は見られなかった．

有意差のあった Q4，Q5 の評点について，要約手法ごとに Freidman 検定を行った結果は，図 3 のとおりである．上述の分散分析と同様に Q4，Q5 の評点においてのみ，統計的有意差（$\chi^2=33.14, p<.01$，$\chi^2=23.70, p<.01$）が見られた．Holm 法による多重比較では，Q4 および Q5 について，提案手法＞重要発言＋反応発言＞重要発言のみ，提案手法＞重要発言＋問いかけ発言＞重要発言のみの関係で統計的有意差があり（$p<.05$），重要発言＋反応発言，重要発言＋問いかけ発言の間では，統計的な有意差は見られなかった（Q4:$p=.46$），Q5:$p=.64$）．

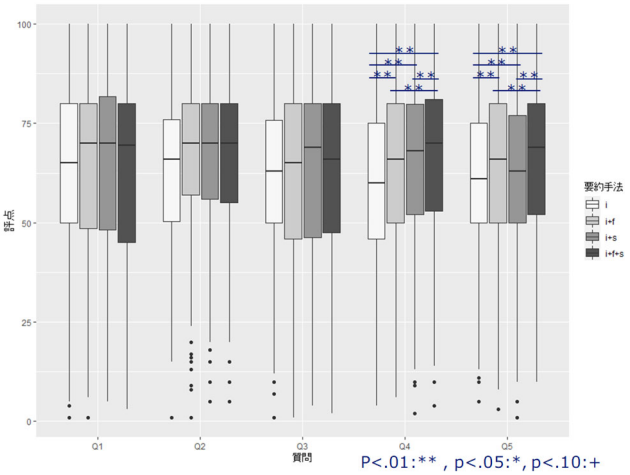

図 3 要約手法の評点の比較（(1)(4)(5)(6)）
i:(重要発言のみ，i+f:重要発言と反応発言，
i+s:重要発言と問いかけ発言，i+f+s:提案手法

5.3 要約による対話印象の変化

各対話の提案手法での要約による，雰囲気の変化は図 4 のとおりである．オリジナルの対話の雰囲気の変化を 0 として，対話ごとに要約によって変化した雰囲気の差分を示している．横軸は雰囲気尺度の各項目（10 項目）であり，縦軸は 9 件法のオリジナル対話からの回答値の差分（マイナス方向に行くほどポジティブ，プラス方向に行くほどネガティブな変化）を示している．比較対象として，重要発言のみで要約した場合の雰囲気の変化を図 5 に示している．

提案する要約手法のバリエーションの比較 5.2 節の結果で確認した通り，提案手法と重要発言のみによる要約手法とでは，提案手法の方が，オリジナル対話より雰囲気の変化が少ない（変化量が 0 に近い）．

提案手法における対話ごとの雰囲気の変化量について比較すると，対話 3，対話 5 について，要約によって大きく雰囲気が変化している．これらの変化が大きな対話についてさらに分析を行うと，他の対話と比較して発言数が少なく，無発言時間が長い，静かな印象を受ける対話であっ

た.

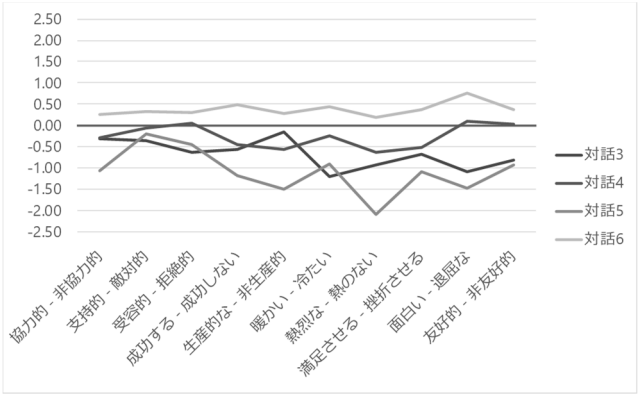

図 4 雰囲気の変化（提案手法）

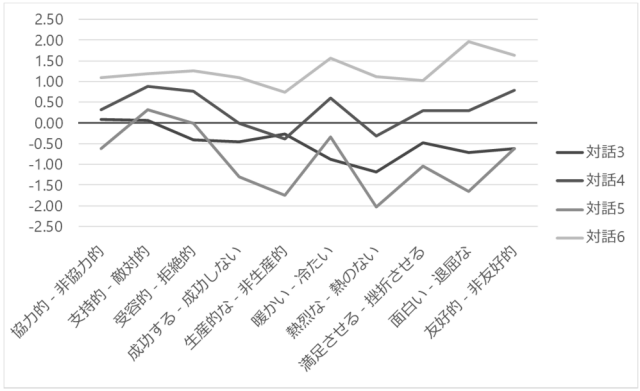

図 5 雰囲気の変化（重要発言のみ）

6. 考察

6.1 提案手法の評価

提案手法と，ランダムに発言を抜き出すランダム要約手法と，専門家による人手による要約手法とを比較した結果，提案手法は比較した内容・雰囲気の再現の全ての観点において、ランダム要約手法より優れていた（仮説 H1-1：支持）．検出力の事後評価では，有意水準 α=.05，効果量 *partial* η^2=.015，f=0.123，サンプルサイズ n=1342 として，検出力(1-β)=.80 であり十分な検出力を持っていた．また，「Q4:雰囲気を示している」「Q5:話者関係を反映している」の評価において，提案手法と人手による要約手法とでは，有意差が見られなかった（仮説 H1-2：一部支持）．十分なサンプルサイズを用いて評価を行っていることを考えると，雰囲気や話者関係の伝達においては，人手による要約とあまり大きな差がないか，同等の評価である可能性が高い．今後，事前の仮定を置いたうえでの実験・同等性検定などで確認する必要がある．

一方で「Q2:内容の反映度」「Q3:納得度」に関しては，人手による要約が優れており，提案手法に改善の余地があると考えられる．今後，人手による要約と提案手法との抽出箇所を比較することで，要約アルゴリズムを改良することが期待できる．

対話の主効果が確認された「Q3:納得度」については，他の評価項目と比較して対話自体の内容に影響を受けやすく，要約手法の是非ではなく，被験者にとって納得感のある議論・結論であるかを評価した可能性がある．特に実験 1 では，被験者ごとの作業時間が短く，オリジナルの対話映像との比較を行うことができない実験設計であったたため，この傾向が強まったと考えられる．

6.2 提案手法のバリエーションの比較評価

重要発言のみを抽出する要約と比較して，対話において意見を求める問いかけ発言や，意見に対して賛成や反対などの態度を示す反応発言をあわせて抽出することは，特に対話の雰囲気や話者関係を伝えるために効果的であることが示された（仮説 H2：一部支持）．

ただし，原理的に，重要発言に問いかけ発言や反応発言を加えると視聴時間が長くなる(問いかけ発言:平均 30 秒，12%増加，反応発言：平均 1 分 28 秒，34%増加，両方：平均 1 分 58 秒，45%増加）ため，対話内容や会全体の雰囲気など,知りたい目的に応じて追加する発言の分量を変える，対話において特に関心があるトピック周辺のみ抽出方法を変えるなどの応用を考える必要がある．

6.3 発言抽出法による雰囲気の変化

重要発言のみによる要約と比較して，提案手法の方が，より対話の雰囲気の変化を抑え，正確に伝える傾向があるといえる（仮説 H3：支持）．しかしながら，個別の対話の雰囲気の変化をみると，対話 5 や対話 3 における熱烈さや生産性など，要約によって大きくポジティブ方向へ変化している項目がある.これらの対話を他の対話と比較すると，全編を通して発言数が少なく，参加者全員で考える時間が長く，特定の参加者に発言が偏る，静かな対話という特徴があった．今回の提案手法では，発言箇所を重要さや談話行為に応じて抽出する要約手法であるため，参加者が考え込むような無発言区間は無視されてしまう．その結果，実際の対話よりも，時間当たりの発言数が多く，活発で生産的な雰囲気を与える要約となった可能性が高い．

これら対話の雰囲気をより正確に伝えるためには，オリジナルの対話の発言数や無発言区間などの特徴に応じて，採用するアルゴリズムを切り替え，場合によっては敢えて無発言区間などを要約に含めるなどの方法が考えられる．

7. おわりに

本研究では，対話映像の要約において，内容だけでなく対話全体や参加メンバの雰囲気を伝える手法を提案し，被験者実験によって有効性を検証した．2 回の評価実験の結果，提案手法は，ベースラインとしてのランダム要約より内容・雰囲気を伝えることができること，専門家による人手の要約と比較して，雰囲気や話者関係の伝達において，同等の表現ができる可能性が示唆された．また，重要度のみの要約と比較して，雰囲気や話者関係の伝達において，優れていること，雰囲気をより正確に伝達できることを示した．

今回の実験では，要約手法の比較評価に重点を置いていることから，重要発言や談話行為などのアノテーションを人手で付与した対話データ（教師データ）を用いて評価を行った．今後，学習済みのモデルおよび異なるデータセットを用いて，対話データの評価を行っていきたい．

参考文献

[1] Feng, X., Feng, X. and Qin, B.. A Survey on Dialogue Summarization: Recent Advances and New Frontiers.
[2] Apostolidis, E., Adamantidou, E., Metsai, I. A., and Patras, I.. Video Summarization Using Deep Neural Networks: A Survey. Proceedings of the IEEE. vol. 109, no. 11, Nov. 2021, p. 1838-1863.
[3] 野中郁次郎, 加賀野忠男, 小松陽一, 奥村昭博, 坂下明宣. 組織現象の理論と測定. 千倉書房 1978.
[4] Rajpoot, V. and Girase, S.. A Study on Application Scenario of Video Summarization. Proceedings of the 2nd International conference on Electronics, Commnunication and Aerospace Technology (ICECA). 2018, p. 936-943.
[5] Gygli, M., Grabner, H., Riemenschneider, H., & Gool, L. V. Creating summaries from user videos. Proceedings of the European Conference on Computer Vision (ECCV 2014). 2014, p. 505-520.
[6] Song, Y., Vallmitjana, J., Stent, A., and Jaimes, A.. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, p. 5179-5187.
[7] Huang, Y., Feng, X., Feng, X., and Qin, B.. The factual inconsistency problem in abstractive text summarization: A survey. arXiv:2104.14839, 2021.
[8] 近藤公久, 釜島萌. 対話者の感情状態による場の空気の推定―対話音声コーパスを用いた検討―. 日本感性工学会論文誌, 2016, vol.15, no. 2, p. 279-285.
[9] 二瓶芙巳雄, 中野有紀子. マルチモーダル情報に基づく重要発言推定モデルを搭載した議論要約ブラウザの有効性の検証. ヒューマンインタフェース学会論文誌, 2020, vol. 22, no. 2, p. 137-150.
[10] ISO 24617-2:2020 “Language resource management − Semantic annotation framework (SemAF) − Part 2: Dialogue acts”. https://www.iso.org/standard/76443.html, (参照 2022-04-11).
[11] Guidelines for Dialogue Act and Addressee Annotation Version 1.0 https://groups.inf.ed.ac.uk/ami/corpus/Guidelines/dialogue_acts_manual_1.0.pdf (参照 2022-04-11).
[12] 林佑樹, 二瓶芙巳雄, 中野有紀子, 黄宏軒, 岡田将吾, グループディスカッションコーパスの構築および性格特性との関連性の分析. 情報処理学会論文誌, 2015, vol.56, no.4, p. 1217-1227.
[13] Nihei, F., Nakano, Y. I., Hayashi, Y., Huang, H., and Okada, S.. Predicting Influential Statements in Group Discussions using Speech and Head Motion Information, 16th ACM International Conference on Multimodal Interaction (ICMI). 2014, p. 136-143.